

Merken, Hans

Evaluation in der Erziehungswissenschaft - eine neue Herausforderung?

Zeitschrift für Pädagogik 50 (2004) 1, S. 77-87



Quellenangabe/ Reference:

Merken, Hans: Evaluation in der Erziehungswissenschaft - eine neue Herausforderung? - In: Zeitschrift für Pädagogik 50 (2004) 1, S. 77-87 - URN: urn:nbn:de:0111-opus-47987 - DOI: 10.25656/01:4798

<https://nbn-resolving.org/urn:nbn:de:0111-opus-47987>

<https://doi.org/10.25656/01:4798>

in Kooperation mit / in cooperation with:

BELTZ

<http://www.beltz.de>

Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document.

This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Kontakt / Contact:

peDOCS
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Mitglied der


Leibniz-Gemeinschaft

Inhaltsverzeichnis

Thementeil: Bildung über die Lebenszeit

Reinhard Fatke

Bildung über die Lebenszeit – Einführung in den Thementeil	1
--	---

Aleida Assmann

Der väterliche Bücherschrank – Über Vergangenheit und Zukunft der Bildung	5
---	---

Rita Casale/Jürgen Oelkers/Daniel Tröhler

Lebenslanges Lernen in historischer Perspektive – Drei Beispiele für ein altes Konzept	21
---	----

Ludwig Liegle/Kurt Lüscher

Das Konzept des „Generationenlernens“	38
---	----

Helmut Fend/Fred Berger/Urs Grob

Langzeitwirkungen von Bildungserfahrungen am Beispiel von Lesen und Computer Literacy – Ergebnisse der LIFE-Studie (Lebensverläufe von der späten Kindheit ins frühe Erwachsenenalter)	56
--	----

Diskussion 1: Evaluation der Erziehungswissenschaft

Hans Merzens

Evaluation in der Erziehungswissenschaft – eine neue Herausforderung?	77
---	----

Heinz-Elmar Tenorth

Wer hat Angst vor den „Evaluationskommissaren“? Evaluation und Qualitätssteigerung der erziehungswissenschaftlichen Forschung	88
--	----

Frieda Heyting

Beobachtungen zur internationalen Anschlussfähigkeit der Allgemeinen Erziehungswissenschaft in Deutschland	99
---	----

Diskussion 2: „Neue Kindheit“ in der Diskussion

Horst Rumpf

Für einen entdramatisierenden Umgang mit der „neuen Kindheit“?

Zur Abwehr pädagogischer Horrorvisionen.

Eine Erwiderung auf Jürgen Oelkers 112

Jürgen Oelkers

Eine Erwiderung auf Horst Rumpf 116

Sammelbesprechung

Hans-Joachim von Olberg

Didaktik auf dem Wege zur Vermittlungswissenschaft?

Eine Sammelbesprechung neuerer Veröffentlichungen 119

Besprechungen

Joachim Kersten

Rudolf Leiprecht: Alltagsrassismus. Eine Untersuchung bei Jugendlichen in
Deutschland und den Niederlanden

Benno Hafenecker/Mechtild M. Jansen: Rechte Cliques. Alltag einer neuen
Jugendkultur

Edith Wölfl: Gewaltbereite Jungen – was kann Erziehung leisten? Anregungen
für eine gender-orientierte Pädagogik 132

Gisela Miller-Kipp

Karl Heinz Jahnke: Jugend unter der NS-Diktatur 1933–1945. Eine Dokumentation

Michael Buddrus: Totale Erziehung für den totalen Krieg. Hitler-Jugend und
nationalsozialistische Jugendpolitik. 2 Bde.

Hermann Langer: „Im gleichen Schritt und Tritt“. Die Geschichte der
Hitler-Jugend in Mecklenburg von den Anfängen bis 1945 138

Alois Suter

Robert Döpp: Jenaplan-Pädagogik im Nationalsozialismus. Ein Beitrag zum
Ende der Eindeutigkeit 142

Dokumentation

Pädagogische Neuerscheinungen 146

Hans Merkens

Evaluation in der Erziehungswissenschaft – eine neue Herausforderung?

Zusammenfassung: In dem Beitrag wird zunächst die zunehmende Bedeutung von Evaluation in den weiteren Kontext des gegenwärtigen Strukturwandels der Hochschullandschaft gestellt. Knappheit der Mittel und neue Möglichkeiten einer auf empirischen Daten basierenden vergleichenden Evaluation führen zu offenen Formen von Konkurrenz und Wettbewerb, wobei nicht immer eine Fairness des Vergleichs gewährleistet ist. Für die Erziehungswissenschaft ergeben sich besondere Probleme, weil die Ausstattung des Faches sowie auch die Lehrbelastung an den einzelnen Standorten sehr unterschiedlich sein kann. Aufgrund des Entwicklungsstandes der Disziplin ergeben sich auch Probleme bei der Anwendung von Kriterien wie Drittmitteleinwerbung und wissenschaftliche Produktivität, gemessen in Qualität und Quantität von Publikationen.

1. Strukturwandel der Hochschulen

Lange Zeit ist die Dynamik der Entwicklung im Hochschulbereich in Deutschland neben dem Anspruch durch Neugründungen Strukturpolitik zu betreiben, wie das in den 70er-Jahren des letzten Jahrhunderts beispielsweise mit den Umwandlungen von Fach- oder Pädagogischen Hochschulen in Gesamthochschulen sowie dem Aufbau neuer Universitäten versucht worden ist, in erster Linie von der Nachfrage der Studierenden getragen worden. Eine erste Korrektur wurde gegenüber der Steuerung durch Nachfrage nach Studienplätzen mit dem Numerus clausus eingeführt, indem in einigen Fächern die Zahl der Studienplätze nach Kriterien wie Kosten von Studienplätzen und gesellschaftlichem Bedarf festgelegt wurde. Mit dieser Maßnahme hat sich ein Umdenken angekündigt, das zwar von Beteiligten, insbesondere Studierenden, immer wieder in Frage gestellt worden ist, dessen Konsequenzen sich aber erst jetzt zeigen: Bei der Festlegung der Anzahl von Studienplätzen in einem bestimmten Fach wird nicht mehr die Nachfrage als alleiniges Kriterium angewendet, sondern zusätzlich werden Kriterien wie die Kosten eines Studienplatzes, Leistungen und gesellschaftlicher Bedarf herangezogen, wenn zu entscheiden versucht wird, ob in einem bestimmten Fach Studienplätze eingerichtet oder gestrichen werden. Diese Umorientierung erfordert im Prinzip bereits Evaluation, weil das Erstellen von Kosten-/Nutzenanalysen entsprechende Daten sowie deren Bewertung zur Bestimmung des Nutzens bzw. der Leistungen voraussetzen.

Eine weitere Neuorientierung, die im Ergebnis ähnliche Folgen hat, zeichnet sich gegenwärtig erst ab, ohne dass ihre Konsequenzen bereits überall bemerkt werden: In Deutschland hat es traditionell im Hochschulbereich eine Unterteilung nach mehr wissenschaftsorientierten und eher praxisorientierten Institutionen gegeben, die sich auch heute noch entlang der Differenz Universitäten – Fachhochschulen nachvollziehen lässt, deren Berechtigung aber längst nicht mehr unumstritten ist. Dabei hat sich als ein Ergebnis herausgestellt, dass die Studienplätze an den Universitäten teurer sind als die an

den Fachhochschulen. Von daher gibt es auf der Seite der Regierungen – Bund wie Länder – seit längerer Zeit ein Bemühen, mehr Studienplätze an den Fachhochschulen einzurichten und parallel dazu Plätze an den Universitäten zu streichen. Diese Bewegung hat u.a. auch dazu geführt, generell zu fragen zu beginnen, welche Ausbildungen an Universitäten angesiedelt sein müssen und welche an Fachhochschulen verlagert werden können. So gibt es beispielsweise den Vorschlag, die Ausbildung in bestimmten Lehrämtern an die Fachhochschulen zu verlagern. Ähnliche Überlegungen lassen sich aber auch in Bezug auf andere Studiengänge entdecken. Sie werden vor allem dann formuliert, wenn es an Universitäten und Fachhochschulen Studiengänge gibt, die sich auf ähnliche Gegenstandsbereiche beziehen, wie das bei den Wirtschaftswissenschaften, den Ingenieurwissenschaften sowie der Sozialpädagogik der Fall ist.

Gemeinsam ist diesen neuen Entwicklungen, dass weder das Wissenschaftssystem selbst es ist, welches Vorgaben für die Universitäten formuliert, noch die Wünsche der Studierenden alleine ausschlaggebend sind, wenn Mittel im Hochschulbereich verteilt werden. Entscheidungen werden vielmehr im politischen System gefällt und allenfalls durch Gremien vorbereitet, die an der Schnittstelle zwischen politischem und Wissenschaftssystem angesiedelt sind, wie z.B. dem Wissenschaftsrat. Dieser Teil der Auseinandersetzung lässt sich noch als eine Variante unter die erste genannte Neuorientierung subsumieren. Sie weist aber einen weiteren Aspekt auf, der daraus resultiert, dass in beiden Typen von Hochschulen ähnliche, um nicht zu sagen identische Studiengänge angeboten werden, wie das unter dem Vorzeichen der Umstellung auf BA/MA sichtbar und nachprüfbar wird: Werden Studiengänge mit ähnlichem Ziel zu sehr unterschiedlichen Kosten angeboten, dann interessiert es auf jeden Fall, ob es nicht eine lohnende Variante sein könnte, die teurere Version einzustellen. Damit muss sich das Hochschulsystem neuen Kriterien stellen, wenn um die knapper werdenden Mittel im Bildungssystem konkurriert wird.

Beiden genannten Neuorientierungen ist gemeinsam, dass bei den entsprechenden Entscheidungen vor allem die Kosten in der Lehre eine dominierende Rolle spielen. Dabei wird eher quantitativ bilanziert, indem dem Medium Geld und der Berechnung von Geldäquivalenten die höchste Priorität gewährt wird. Unter dem Aspekt der Qualität ist aber ebenso von Interesse, ob die Qualität der Lehre vergleichbar ist. Dazu wird beispielsweise erhoben, ob Leistungen in der Lehre mit der erwarteten Qualität erbracht werden. Als ein Kriterium kann dabei z.B. gelten, ob die erfolgreichen Absolventinnen und Absolventen eines Studiengangs ihrer Ausbildung entsprechende Positionen im Beschäftigungssystem erhalten. Ein anderes Kriterium kann der Prozentsatz der Studienabbrecher sein, der möglichst niedrig ausfallen soll. Mit Evaluationen, das ist der Anspruch, lässt sich sowohl die quantitative als auch die qualitative Dimension der Bewertung in die Entscheidungsprozesse einbeziehen. Die vorher üblichen Steuerungsinstrumente waren eher rein quantitativ orientiert.

Als erste Erkenntnis bezüglich der Durchführung von Evaluationen lässt sich demnach festhalten, dass sie in aller Regel der Vorbereitung oder Begründung von Entscheidungen in den Hochschulen bzw. den einzelnen Fächern dienen. Genau genommen sollen mit Evaluationen Begründungen für Entscheidungen geschaffen werden, die sich

anders nur schwer oder gar nicht durchsetzen lassen. So wird gegenwärtig in Hochschulen einerseits die Zahl der Studienplätze verringert, weil die staatlichen Subventionen für das Hochschulsystem gekürzt werden. Andererseits wird nicht nur innerhalb der einzelnen Hochschule, sondern verstärkt auch zwischen Fachhochschulen und Universitäten sowie Hochschulen des gleichen Typs innerhalb eines Bundeslandes um Mittel konkurriert. Das ist eine Entwicklung, die sich auch international beobachten lässt. In diesem Wettbewerb kann man sich weniger auf Traditionen berufen, sondern es wird verstärkt nach der aktuellen Leistungsfähigkeit gefragt. Die Evaluation stellt einen Versuch dar, in diesem Entscheidungsprozess Antworten zu finden, die evidenzbasiert sind und es gestatten, für unterschiedliche Informationen einen gemeinsamen Kontext zu schaffen, um politische bzw. bildungspolitische Entscheidungen zu legitimieren. Damit ist die Erwartungshaltung formuliert, aus der heraus die Evaluation durchgeführt wird.

Setzt sich die Entwicklung, dass die Kostenfrage in den Vordergrund tritt, durch, dann treten innerhalb des Hochschulsystems Fragen wie die nach der Differenz zwischen Fächern und Standorten, den Kosten der Differenz, der Notwendigkeit der Differenz sowie der Berechtigung der jetzigen Verteilung der Mittel in den Vordergrund. Diese Entwicklung des Denkens lässt sich allgemein beobachten (vgl. Dierkes/Merken 2002), sie dominiert beispielsweise die Expertisen, die von der OECD für das Bildungssystem erstellt werden (OECD 2001). Sie impliziert jedoch weitergehende Fragen. Die erste und wichtigste lautet: stehen in den verschiedenen Hochschulen, insbesondere aber den beiden Typen von Hochschulen, Aufwand und Ertrag in einem angemessenen Verhältnis zueinander. Damit gerät die einzelne Hochschule – aber auch jedes Fach innerhalb einer Hochschule – unter Rechtfertigungsdruck. Die Antwort auf diese neuen Herausforderungen wird innerhalb der Hochschulen deutlich, wenn Zielvereinbarungen geschlossen werden und ein Controlling eingerichtet wird. Auf diese Weise werden Geldäquivalente als Entscheidungsgrundlage eingeführt.

Eine dritte Neuorientierung kündigt sich dadurch an, dass sich die Annahme, die einzelnen Fächer seien an den verschiedenen Hochschulen jeweils gleich gut vertreten, nicht mehr aufrecht erhalten lässt (Dierkes/Merken 2002). Das ist eine der zentralen Überlegungen, die sich bei den Rankings der verschiedensten Form immer wieder durchgesetzt hat. Hier handelt es sich um ein Denken, das schon lange typisch für die USA ist, an das man sich in Deutschland aber erst allmählich gewöhnt. Rankings bedürfen wiederum in vielen Fällen vorangehend einer Evaluation, weil man erst auf diese Weise die erforderlichen Daten gewinnen kann.

Die bisher beschriebenen Veränderungen und Entwicklungen haben sich innerhalb des gesamten Hochschulsystems ergeben. Sie haben sich auch auf die Situation der Erziehungswissenschaft an den Hochschulen ausgewirkt, die fast unbemerkt eine der Spitzenpositionen in der Gruppe der Disziplinen mit den meisten Studienplätzen erreicht hat (Otto u.a. 2000; Merken/Rauschenbach/Weishaupt 2002). Als Disziplin hat sie lange Zeit von den Veränderungen im Hochschulsystem, insbesondere von der beschriebenen Nachfrageorientierung bei der Einrichtung von Studienplätzen, profitiert, wenn man die Veränderungen in den Stellenzahlen bilanziert. Für die Disziplin stellt die beschriebene neue Basis für Begründungen von Entscheidungen im Hochschulsystem al-

leine schon deshalb eine Herausforderung dar, weil sie ihre gewohnten Muster des Argumentierens innerhalb des Systems, wenn es um die Auseinandersetzung um Ressourcen geht, auf die neuen Entwicklungen ausrichten muss. Der alleinige Verweis auf die Nachfrage nach Studienplätzen reicht nicht mehr aus.

2. Die neuen Herausforderungen durch Evaluation

In Zeiten knapper finanzieller Mittel hat ein verstärkter Wettbewerb zwischen und innerhalb der Hochschulen eingesetzt. Galt es bisher, dass mit anderen Fächern innerhalb der Universitäten neben den Zahlen von Studierenden vielleicht über Erfolge in der Forschung, über Macht in den akademischen Gremien und mittels Beziehungen zu den verantwortlichen Ministerien um Ressourcen konkurriert wird, so wurde nunmehr eine neue Messlatte in Form der Evaluation eingeführt. Mit dieser Messlatte wird beansprucht, sowohl für intrauniversitäre Vergleichsprozesse mit anderen Fächern als auch für interuniversitäre Vergleiche innerhalb des eigenen Faches ein „objektives“ und vor allem transparentes Instrumentarium präsentieren zu können. Dabei lautet der Anspruch vor allem, die Vergleiche mittels „objektiver“ Daten abzusichern. Damit ist eine Praxis in Deutschland implementiert worden, die in anderen Ländern längst bekannt war. Die Niederlande und die USA seien als Beispiele genannt: Das einzelne Fach muss sich am jeweiligen Standort aber auch im Vergleich der Standorte als konkurrenzfähig erweisen, wenn es das eigene Überleben sichern will. Ähnlich stehen die Universitäten im Wettbewerb, für die es Rankings gibt.

Solche Vergleiche erfordern einen Korpus von Kriterien, die auf unterschiedliche Standorte und Fächer angewendet werden können und aus der Perspektive des Hochschulsystems fair sind. Damit ist ein wesentlicher Punkt benannt, der für die Evaluation von Bedeutung ist und oben mit dem Begriff „evidenzbasiert“ benannt worden ist. Evidenzbasierung meint im Kern, dass es sich um ein Bündel von Kriterien handeln muss, das plausibel ist, dessen Anwendung von den Betroffenen also möglichst wenig bestritten werden kann. Die Bewertung eines Standortes oder eines Faches muss im Vergleich zu anderen auf eine Datenbasis gestellt werden, die diese Bewertung als Ergebnis der Evaluation nachvollziehbar macht. Erste Erfahrungen mit Evaluationen diesen Typs konnten Fächer in Universitäten sammeln, wenn in Zeitschriften wie Stern oder dem Spiegel Rankings auf der Basis der Befragung von Studierenden durchgeführt wurden, bei denen dann Universitäten, die weniger nachgefragt wurden, besser abschnitten, weil bei ihnen bestimmte Relationen wie z.B. diejenigen zwischen Professoren und Studierenden, günstiger ausfielen, was zu vereinfachten Kontakten führte. Ebenso interessierte die Anzahl von Arbeitsplätzen in Bibliotheken in Relation zur Zahl der Studierenden, der Mietpreis für ein Zimmer etc. Diese Beispiele belegen, wie sich mit der neuen Praxis des Evaluierens vor allem die Wahl von Kriterien verändert hat. Das ist wichtig, weil Evaluation Messen und Bewerten bedeutet.

Sowohl für die Konstruktion des Maßstabs als auch für den Vorgang des Bewertens bedarf es bei der Evaluation bestimmter Kriterien. Diese müssen vor Beginn der Evalua-

tion festgelegt sein. Für die Zwecke der Evaluation sind allerdings Kriterien, die in der Vergangenheit bereits angewendet worden waren, neu präzisiert worden. Ebenso sind Gewichtungen bereits bekannter Kriterien verändert worden. Es handelt sich beispielsweise jeweils um ein anderes Kriterium, wenn entweder die Zahl der Studierenden, oder die Zahl der Studierenden innerhalb bestimmter Semester oder die Zahl der Studierenden, die innerhalb einer bestimmten Anzahl von Semestern erfolgreich ihr Studium beendet haben, oder die Zahl der Studierenden, die nach einem Examen eine berufliche Tätigkeit aufgenommen haben, als Grunddatum gewählt wird – um ein konkretes Beispiel zu nennen. Ebenso ist es ein Unterschied, ob die Anzahl der Beiträge in „internationalen referierten Zeitschriften, in nationalen peer-reviewten Zeitschriften oder Institutsberichten als Publikationen berücksichtigt werden. „Objektiv“ in dem Sinne, dass der Zahlenwert feststellbar ist, sind alle diese Daten. Wer sich allerdings in der Vergangenheit auf andere Kriterien eingelassen hatte als diejenigen, die gegenwärtig angewendet werden, muss sich relativ kurzfristig umstellen. Das verleiht dem Prozess der Evaluation sowie den Konsequenzen, die aus ihm auf der Seite der Ministerien, der Hochschulen und der Personen jeweils gezogen werden, eine eigene Dynamik.

Mit der Praxis des Evaluierens hat sich zusätzlich verändert, dass sich diejenigen, die evaluiert werden, an der Festlegung der Kriterien kaum noch beteiligen können. Früher sind sie entweder gefragt worden oder haben Einfluss nehmen können. Heute werden die Kriterien von den Kommissionen vorgegeben, die mit der Evaluation beauftragt sind. Das erzeugt bei den von der Evaluation Betroffenen schnell ein Gefühl des Ausgeliefertseins bzw. des willkürlichen Vorgehens seitens der Evaluatoren. Wer beispielsweise in der Vergangenheit darauf gesetzt hatte, über möglichst viele Studierende unabhängig von der Semesterzahl, die sie bereits an der Hochschule verbracht hatten, zu verfügen, wird es als ein unangemessenes Kriterium begreifen, wenn nunmehr nur noch die Studierenden in der Regelstudienzeit positiv einbezogen werden und für Studierende jenseits der Regelstudienzeit möglicherweise gar ein Malus vergeben wird.

„Erfolgsdefinitionen“ bei Studienzeiten, Studienabschlüssen und Übertrittsquoten in das Beschäftigungssystem, das macht ihre Beliebtheit bei Evaluatoren aus, lassen sich studienfachunabhängig bestimmen. Damit ist zunächst auch der Vergleich unterschiedlicher Fächer möglich geworden. Ebenso lässt sich ein anderes Datum – die Kosten des einzelnen Studienplatzes als Vergleichszahl für verschiedene Fächer – ermitteln. Hochschulsysteme müssen dann auf der Basis dieser und ähnlicher Daten für sich eine Planung vornehmen, welche Fächer mit wie vielen Studienplätzen an ihnen vertreten sein sollen. Damit obliegt ihnen eine Steuerungsaufgabe, für die sie bisher wenig vorbereitet sind und die starke Elemente des Controlling in Unternehmen aufweist: Man setzt sich Ziele und muss möglichst zeitnahe Informationen darüber zur Verfügung haben, wie der jeweilige Ist-Stand den Vorgaben entspricht, um Veränderungen vornehmen oder gegensteuern zu können. Gerade die formative Evaluation weist große Ähnlichkeiten mit dem Controlling auf. An die Stelle von politischen Entscheidungen treten Managemententscheidungen, die rational basiert zu sein scheinen, wenn man sich auf den Korpus der Kriterien erst einmal geeignet hat. Dieser Korpus wird wiederum in vielen Fällen von den verantwortlichen Ministerien vorgegeben.

3. Konkurrenz und Wettbewerb als neue Grundfigur des Handelns

Konkurrenz und Wettbewerb gab es in der Vergangenheit innerhalb des Hochschulsystems vor allem dann, wenn es um das individuelle Ansehen von Hochschullehrern gegangen ist. Zumindest an der Oberfläche spielten sie keine Rolle zwischen Hochschulen und innerhalb der Hochschule. Nunmehr werden die Hochschulen mit Rankings konfrontiert und innerhalb der Hochschulen werden auch Rankings zwischen den Fächern erstellt. Damit wird auf der einen Seite ein Wissen transparent gemacht, das es bei Insidern vielleicht schon immer gegeben hat. Dieses Wissen wird zusätzlich im Gegensatz zur Vergangenheit auch empirisch begründet. Zugleich wird auf diese Weise eine Form des Wettbewerbs implementiert, die nicht immer gerecht ist. Ein kleines Beispiel mag das erläutern: Bei dem Bemühen um Drittmittel spielt die Ausstattung der Fächer mit Stellen für wissenschaftliche Mitarbeiter eine große Rolle. Beim Vergleich der Fächer miteinander werden aber oft nur die eingeworbenen Drittmittel zueinander in Relation gesetzt. Die Frage der Gleichheit der Chancen wird kaum gestellt. Ungleichgewichte dieser Art müssten aber in eine faire Bewertung einbezogen werden.

Beim Wettbewerb zwischen verschiedenen Standorten lassen sich ähnliche Verzerrungen beobachten. Häufig gibt es an bestimmten Standorten zwar wissenschaftliche Mitarbeiter. Wenn diese aber überwiegend in Dauerstellen beschäftigt sind, stellen sie für das Einwerben von Drittmitteln auf der einen Seite keine Ressource dar und können auf der anderen Seite dazu beitragen, dass die Prüfungsbelastung der Hochschullehrer z.B. bei den Prüfungen in den Lehramtsstudiengängen sehr stark anwächst (Weishaup/Merkens 2000). Damit wird ein Problem angesprochen, das in der Praxis der bisherigen Evaluationen keine Rolle gespielt hat, dem aber entscheidende Bedeutung zukommt, wenn die Relation zwischen Aufwand und Ertrag fair bestimmt werden soll: Die Leistungsfähigkeit eines Systems lässt sich eben auch dadurch ermitteln, dass die Relation zwischen Input und Output bestimmt wird und nicht nur der Output in die Bilanzierung einbezogen wird. In diesem Sinne sind viele der Evaluationen, welche es bisher gibt, rein auf den Output bezogen. Das muss im Ergebnis zu Ungerechtigkeiten führen wie der, dass Universitäten mit einer höheren Attraktivität für Studierende und hohen Zahlen an Studierenden bei der Betreuungsintensität schlechter abschneiden als Universitäten, die wenig attraktiv sind. Spätestens an dieser Stelle wird ersichtlich, dass die Ergebnisse von Evaluationen unter Umständen mehr Fragen aufwerfen können, als sie Antworten zu geben scheinen. Evaluationen können den Erwartungen, die an sie gestellt werden, erst dann gerecht werden, wenn sie Input und Output sowie die dazwischenliegenden Prozesse in die Erhebungen einbeziehen, die in den Hochschulen stattfinden, um den Input in den Output zu transformieren. Während wir bisher eher eine produktbezogene Qualitätskontrolle als wesentlichen Teil der Evaluation praktizieren benötigen wir in Ergänzung eine prozessbezogene Qualitätskontrolle.

Wenn Evaluationen zur Verbesserung der Leistungsfähigkeit des Wissenschaftssystems insgesamt und auch der Erziehungswissenschaft im Besonderen beitragen sollen, dann reicht die Form der summativen Evaluation, die häufig gewählt wird, nicht aus. Es muss vielmehr formativ evaluiert werden (Scriven 1967). Erst die formative Evaluation

gestattet es, den Prozess der Leistungserbringung einzubeziehen. In dem Verzicht darauf muss ein Manko der bisherigen Praxis des Evaluierens gesehen werden.

Um die Steuerung des Hochschulsystems und der einzelnen Hochschulen zu verbessern wird es in der Zukunft demnach eines neuartigen Instrumentariums bedürfen, das sich in der Praxis des bisherigen Evaluierens allenfalls in Ansätzen abzeichnet: Man muss die Leistungserbringung in den Hochschulen zeitnah dokumentieren. Das kann mit der Umstellung auf Bachelor und Master schon eher gelingen, weil das Prinzip der gestuften Prüfungen Fortschritte der Studierenden im Studium leichter im Verlauf feststellen lässt. Allerdings wird der Aufwand für Dokumentation gleichzeitig erheblich zunehmen.

4. Das Grundschema zwischen Evaluator und Evaluierten

Ein Dissens zwischen Evaluierten und Evaluatoren resultiert aus einer unterschiedlichen Interessenlage: Die Kriterien, welche die Evaluatoren anwenden, sind in der Regel nicht identisch mit den Kriterien, welche die Evaluierten für die Bewertung der eigenen Arbeit als wichtig ansehen. Dieser Dissens kann zwar abgemildert werden, ist aber prinzipiell nicht überwindbar. Die Aktionsforschung hat ihn zu negieren versucht, indem die Evaluatoren zu Handelnden und die Handelnden zu Evaluatoren mutierten. Resultat war ein doppelter Dilettantismus, weil die Evaluatoren nicht immer über die Kompetenzen der im Feld Handelnden und die im Feld Handelnden nicht über die Kompetenzen der Evaluatoren verfügt haben. Für den Gewinn an kommunikativer Validität, der mit diesem Modell gerne verknüpft worden ist, ist also ein hoher Preis bezahlt worden. Jedoch ist nicht zu übersehen, dass dem Ansatz der Aktionsforschung bei Evaluationen im Wissenschaftssystem prima facie eine hohe Plausibilität zugebilligt werden kann, weil die wechselnden Positionen der Protagonisten in Bezug auf Evaluator und Evaluierter solche Lösungen geradezu nahe legen. Jedoch darf nicht übersehen werden, dass es hier nicht nur im Einzelfall zu Differenzen kommen muss. So gibt es in der Erziehungswissenschaft noch immer einen Forschungstypus, in dessen Rahmen ein Wissenschaftler ein Opus magnum vorlegt, das auf personengebundener Einzelforschung ohne Einsatz von Drittmitteln basiert. Aus dieser Sicht kommen Buchpublikationen eine überragende Bedeutung zu. Es muss allerdings Buch für Buch überprüft werden, welche Qualität ihm zukommt. Das kollidiert mit den Gepflogenheiten der Evaluatoren, leicht zugängliche Kriterien zu verwenden. Dass dieser Weg Beschwerden verursacht, belegen die Standards, die die Deutsche Gesellschaft für Evaluation (2002) vorgelegt hat. Danach müssen Evaluationen den Standards der Nützlichkeit, Durchführbarkeit, Fairness und Genauigkeit genügen. Insbesondere die Nützlichkeit und die Fairness werden von den Evaluierten häufig in Zweifel gezogen werden. Das resultiert auch in der Erziehungswissenschaft vor allem aus einem vermuteten Zweck, der häufig mit der Evaluation verbunden wird: Es geht häufig darum, Gründe für geplante Einsparungen bereitzustellen. Dabei wird das Ziel häufig auch von denen nicht benannt, die die Evaluation in Auftrag gegeben haben. Das müsste aber gemäß den Standards der Deutschen Gesell-

schaft für Evaluation unter dem Aspekt der Nützlichkeit geleistet werden. Wenn vorgegebener (Leistungsvergleiche) und heimlicher Zweck der Evaluation (Einsparpotenziale im Wissenschaftsbereich) weit auseinander klaffen, dann resultiert daraus eine Glaubwürdigkeitsfalle.

5. Das Fach Erziehungswissenschaft

Für das Fach Erziehungswissenschaft ergeben sich bei der Evaluation Probleme, die für andere Fächer in dieser Weise nicht typisch sind. Zunächst ist bemerkenswert, dass in der Lehre erhebliche Teile des Lehrdeputats für das Begleitstudium in der Lehrerbildung erbracht werden müssen. An einzelnen Standorten werden deutlich mehr als 50% des vorhandenen Lehrdeputats in diesem Bereich abgerufen. Das führt im Einzelfall dazu, dass die Erziehungswissenschaft zwar in Lehre und Forschung, aber nur in ihrer Funktion in der Lehrerbildung evaluiert worden ist (Wissenschaftliche Kommission Niedersachsen 2002). Dabei handelt es sich bei dem erziehungswissenschaftlichen Anteil – bezogen auf das Gesamtvolumen der Semesterwochenstunden der Studierenden in der Lehrerbildung – jeweils nur um einen kleinen Anteil. Daraus resultiert, dass ein deutliches Profil für das Fach aus diesem Teil der Lehre nur schwer gewonnen werden kann, obwohl es heute Bestrebungen gibt, diesen Anteil und auch die Erziehungswissenschaft auf Bildungsforschung zu fokussieren (Weiler 2003). Damit ist eine Diskussion von außen an die Disziplin herangetragen worden; sie selbst hat diese Eingrenzung nicht vorgeschlagen.

Wenn das Lehrdeputat in einem so großen Umfang für eine bestimmte Aufgabe aufgebracht werden muss, ohne dass daraus eine Profilbestimmung für die Disziplin als ganze folgen kann, stellt sich wie von selbst die Frage nach der Mindestausstattung, die von anderen Disziplinen, wie vor allem der Psychologie, immer an den Beginn gestellt wird, die aber in der Erziehungswissenschaft bisher nur bei dem Neuaufbau in den neuen Bundesländern als Forderung für die Ausstattung der Standorte formuliert worden ist. Eine Mindestausstattung muss sowohl für die Begleitstudien in der Lehrerbildung als auch für die Hauptstudiengänge in Diplom und Magister bzw. zukünftig BA und MA bestimmt werden. Das gilt sowohl hinsichtlich der Anzahl der benötigten Stellen als auch hinsichtlich der Denomination dieser Stellen. Es ist bezeichnend für die Situation in der Erziehungswissenschaft, dass diese Frage bisher kaum diskutiert und schon gar nicht gelöst worden ist. In den Evaluationsberichten, die bisher erstellt worden sind, findet sie sich auch nicht. Es gibt allenfalls implizite Hinweise darauf, dass die Evaluatoren Überlegungen in diese Richtung mit berücksichtigt haben, wenn sich Vorschläge für bestimmte Ergänzungen des Spektrums an einzelnen Universitäten finden. Zu fragen wäre aber zuerst, welches Stundendeputat in welchem Studiengang bei welcher Studien- und Prüfungsordnung erforderlich ist, um dem gesetzlichen Auftrag Folge leisten zu können, dass die Lehre ordnungsgemäß erbracht wird.

Die Bestimmung der Mindestausstattung müsste am Beginn eines Berichtes geleistet werden, weil sich die Leistungen, die erzielt werden, nur in Relation zu dem erfassen lassen, was an Ausstattung vorhanden ist. Dabei kann eine quantitative von einer qualitati-

ven Fixierung unterschieden werden. Während die quantitative Fixierung verhältnismäßig einfach fällt, gibt es bei der qualitativen in der Erziehungswissenschaft erhebliche Schwierigkeiten, weil es bisher noch keinen Konsens darüber gibt, was das Kerncurriculum der Erziehungswissenschaft ist bzw. wie das Kerncurriculum in den verschiedenen Studiengängen bestimmt ist, an denen die Erziehungswissenschaft in den Universitäten beteiligt ist. Das ist aber eine wesentliche Voraussetzung dafür, um die Zahl der benötigten Stellen mit Denominationen für diese Stellen zu versehen. Dies wiederum hat zur Folge, dass auch die Evaluatoren Probleme haben werden, einerseits die erbrachten Leistungen zu bewerten und andererseits Vorschläge zu entwickeln, welche Veränderungen vorgenommen werden müssten.

In einem zweiten Bereich, der für die Evaluationen zentral ist, gibt es ebenfalls erhebliche Probleme, wenn fair evaluiert werden soll. Diese Probleme haben sich noch dadurch verstärkt, dass in der letzten Zeit der Anspruch immer stärker wird, Wissenschaftlerinnen und Wissenschaftler aus anderen Ländern für die Mitarbeit bei den Evaluationen zu gewinnen: In den Niederlanden, Belgien und Großbritannien ist es beispielsweise in der Erziehungswissenschaft, aber auch in anderen Disziplinen längst üblich geworden, bei der Forschung nicht nur den Input, das sind die eingeworbenen Drittmittel, sondern ebenso den Output in die Evaluation mit einzubeziehen. Bei Letzterem ist dann ein wesentliches Kriterium für die Bewertung des Erfolges, welche Teile der scientific community vom Output unterrichtet werden konnten. Um das feststellen zu können, gibt es ein Standardinstrumentarium: Es wird nach den Organen gefragt, in denen Ergebnisse publik gemacht worden sind. Dann kann zwischen internationalen Zeitschriften mit einem hohen impact factor, englischsprachigen oder anderen fremdsprachlichen Publikationen, nach peer-reviewten nationalen Zeitschriften, anderen Zeitschriften, Buchpublikationen und grauer Literatur unterschieden werden. Diese Unterscheidung wird auch in den entsprechenden Ländern praktiziert, wenn dort evaluiert wird. Es gibt ähnliche Unterscheidungen in Deutschland in den Naturwissenschaften, der Medizin und den technischen Disziplinen, sie bereiten auch den so genannten kleineren Fächern keine Probleme. In der Erziehungswissenschaft wird die Situation aber prekär, wenn solche Indikatoren angewendet werden. Das beginnt bereits, wenn die Frage nach den peer-reviewten nationalen Zeitschriften gestellt wird. Hier verfügen wir in der Bundesrepublik über ein viel zu geringes Aufkommen. Das hat dann auch zur Folge, dass zu wenige Zeitschriften in den social science citation index aufgenommen worden sind, der ebenfalls als Quelle herangezogen werden könnte, wenn das wissenschaftliche Ansehen geprüft werden soll. Damit verfügt die Erziehungswissenschaft als Disziplin gerade bei dem Wettbewerb mit anderen Disziplinen der Geistes- und Sozialwissenschaften über schlechte Ressourcen in einem zentralen Bereich: Wenn es um die disziplinäre Sichtbarkeit als Bewertungskriterium für Erfolg oder Misserfolg geht, bietet die Disziplin keine guten Voraussetzungen. Daraus resultiert, dass die Erziehungswissenschaft im inneruniversitären Vergleich mit anderen Disziplinen häufig Schwierigkeiten hat, die eigene Leistungsfähigkeit darzustellen.

Neben dem in der Lehre markierten Manko zeichnet sich demnach auch ein Problem ab, wenn es um die Darstellung und Bewertung der Forschungsleistungen geht. In

diesem Bereich lässt sich auch noch eine weitere Schwierigkeit verorten: Viele der Drittmittel, die eingeworben werden, stammen von Ministerien oder Verbänden. Es handelt sich dabei in aller Regel nicht um Mittel, für deren Vergabe Anträge peer-reviewed worden sind. Anders als in vielen technischen und naturwissenschaftlichen Disziplinen, in denen in großem Umfang Mittel ebenfalls von Dritten eingeworben werden, die an Ergebnissen angewandter Forschung interessiert sind und kein umfangreiches Reviewsystem eingerichtet haben, verfügt die Erziehungswissenschaft über verhältnismäßig geringe Mittel, die bei der DFG oder ähnlichen Stiftungen eingeworben worden sind. Die prestigeträchtigen Mittel werden in der Erziehungswissenschaft weniger nachgefragt. Daraus resultiert eine eigenartige Situation: Obwohl das Drittmittelaufkommen der Erziehungswissenschaft hoch ist, hat das nicht zu einer entsprechenden Anhebung des Ansehens der Disziplin geführt.

Wenn man die bisherigen Evaluationen bilanziert, von denen Erziehungswissenschaft mitbetroffen war, oder die nur für Teile von ihr durchgeführt worden sind, dann gibt es ein Manko, sobald man die Evaluatoren betrachtet, soweit diese aus Deutschland kommen. In vielen Fällen ist es ihnen nicht gelungen, auf die verschiedenen Fälle, die zur Evaluation anstanden, einheitliche Kriterien anzuwenden. Es sind vielmehr in der Praxis häufig von Hochschule zu Hochschule Kriterien verändert worden. Das führt zu misslichen Resultaten, weil am Ende eines solchen Vorgehens nicht mehr vermittelt werden kann, warum das, was bei der Hochschule A moniert wurde, bei einer anderen nicht berücksichtigt worden ist etc. Was hier beklagt wird, ist die mangelnde Professionalität vieler Evaluatoren im internationalen Vergleich. Peer-review-Verfahren zeichnen sich dadurch aus, dass diejenigen, die evaluieren, in einem anderen Verfahren zu den Evaluierten zählen können. Das ist der Vorteil dieses Systems. Sein Funktionieren setzt aber auf Dauer voraus, dass es bestimmte Standards gibt, die im Verfahren zu beachten sind. Die Deutsche Gesellschaft für Evaluationsforschung hat solche Standards formuliert. Ihre Beachtung sollte eine *conditio sine qua non* in entsprechenden Verfahren sein. Das ist bisher leider in vielen Fällen nicht gegeben.

6. Fazit: Evaluation in der Erziehungswissenschaft – was muss sich ändern?

Das Fach Erziehungswissenschaft wird sich der Mode der Evaluation nicht entziehen können. Damit es hier zu angemesseneren Resultaten kommt, ist eine Änderung bestimmter Randbedingungen erforderlich. Dazu kann die Disziplin Vieles beitragen.

1. Ein wichtiges Ziel muss es sein, ein *Kerncurriculum* der Disziplin und ihrer Teildisziplinen zu entwickeln, das als verbindlich angesehen werden kann. Erst auf der Basis eines solchen gemeinsamen Kerncurriculums können Besonderheiten einzelner Standorte abgebildet und weiterentwickelt werden.
2. Ebenso ist es erforderlich, die Frage der *Mindestausstattung* an den einzelnen Standorten – unter Berücksichtigung der jeweils vorhandenen Studiengänge – zu bestimmen. Die bisherige Praxis, mit minimalster Ausstattung mehrere Studiengänge

anzubieten, kann im Endeffekt nicht zu befriedigenden Resultaten der Evaluation führen.

3. Die *Kriterien*, die bei der Evaluation anderer und vor allem von Nachbardisziplinen aus den Geistes- und Sozialwissenschaften angewendet werden, müssen dahingehend überprüft werden, wie weit sie auch für die Evaluation der Erziehungswissenschaft tauglich sind. Stellt sich dabei heraus, dass das nicht der Fall ist, muss versucht werden, die Entwicklung in der Erziehungswissenschaft dahingehend voranzutreiben, dass solche Kriterien angewendet werden können. Ein Beispiel hierfür ist, dass die Disziplin mehr Zeitschriften mit einem strengen Gutachterverfahren benötigt.
4. Evaluatoren aus der Erziehungswissenschaft müssen sich bemühen, für ihre Arbeit und die Ergebnisse dieser Art das erforderliche Maß an *Transparenz* zu sichern.

Literatur

- Deutsche Gesellschaft für Evaluation (2002): Standards für Evaluation. Köln: Deutsche Gesellschaft für Evaluation.
- Dierkes, M./Merkens, H. (2002): Zur Wettbewerbsfähigkeit des Hochschulsystems in Deutschland. Berlin: Wissenschaftszentrum Berlin WZB.
- Merkens, H./Rauschenbach, T./Weishaupt, H. (Hrsg.) (2002): Datenreport Erziehungswissenschaft 2. Opladen: Leske + Budrich.
- Otto, H.-U. et al (2000): Datenreport Erziehungswissenschaft. Opladen: Leske + Budrich.
- Weishaupt, H./Merkens, H. (2000): Forschung und wissenschaftlicher Nachwuchs. In: Otto u.a. (Hrsg.): Datenreport Erziehungswissenschaft. Opladen: Leske + Budrich, S. 117-134.
- OECD (2001): Bildung auf einen Blick. OECD-Indikatoren. Paris: OECD.
- Scriven, M. (1967): The Methodology of Evaluation. In: Tyler, R./Gagne, R./Scriven, M. (Hrsg.): Perspectives of Curriculum Evaluation. AERA Monograph Series on Curriculum Evaluation, Bd. 1. Chicago: Rand McNally, S. 39-83.
- Weiler, H. (2003): Bildungsforschung und Bildungsreform: Von den Defiziten der westdeutschen Erziehungswissenschaft. In: Gogolin, I./Tippelt, R. (Hrsg.): Innovation durch Bildung. Beiträge zum 18. Kongress der DGfE (München 2002). Opladen: Leske & Budrich, S. 181-203.

Abstract: *The author begins by placing the growing importance of evaluation into the broader context of today's structural changes in the landscape of higher education. The shortage of funds and new possibilities of comparative evaluation based on empirical data lead to open forms of competition in which fairness of comparisons is not always guaranteed. With regard to educational science, specific problems arise because both the available resources and the teaching load at the individual sites may differ considerably. Due to the discipline's level of development there may also arise problems in the application of criteria such as the acquisition of funds from third parties or scientific productivity, assessed by the quality and quantity of publications.*

Anschrift des Autors:

Prof. Dr. Hans Merken, Institut für empirische und vergleichende Erziehungswissenschaft, Freie Universität Berlin, Fabeckstr. 13, 14195 Berlin.